Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
Journal of Multivariate Analysis 85 (2003) 234–252
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

Dimension reduction in partly linear error-in-response models with validation data ☆

Qihua Wang

Academy of Mathematics and System Science, Chinese Academy of Science, Beijing 100871, China and Heilongjiang University, Harbin 150080, China

Received 22 April 1999

Abstract

Consider partial linear models of the form $Y = X^{\tau}\beta + g(T) + e$ with Y measured with error and both p-variate explanatory X and T measured exactly. Let $\tilde{Y}$ be the surrogate variable for Y with measurement error. Let primary data set be that containing independent observations on $(\tilde{Y}, X, T)$ and the validation data set be that containing independent observations on $(Y, \tilde{Y}, X, T)$, where the exact observations on Y may be obtained by some expensive or difficult procedures for only a small subset of subjects enrolled in the study. In this paper, without specifying any structure equations and distribution assumption of Y given $\tilde{Y}$, a semiparametric dimension reduction technique is employed to obtain estimators of β and $g(\cdot)$ based the least squared method and kernel method with the primary data and validation data. The proposed estimators of β are proved to be asymptotically normal, and the estimator for $g(\cdot)$ is proved to be weakly consistent with an optimal convergent rate.

AMS 1991 subject classifications: primary 62J02; secondary 62G05; 62E20

Keywords: Dimension reduction; Partial linear model; Validation data; Asymptotic normality

1. Introduction

Many variables of interest are difficult or expensive to be measured accurately and hence are usually replaced by surrogate observations, which are available by some relatively simple measuring methods. Generally, the relationship between the surrogate variables and the true variables can be rather complicated compared to

☆ Research supported by the NNSF of China.
E-mail address: qhwang@amss.ac.cn.

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(02)00066-0

the classical additive error structure usually assumed [7]. Actually, in many practical settings, it is even difficult to specify the relationship between true variables and their surrogates. That is, the most realistic situation may be that no error structure or distribution assumption of true variables given the surrogate variables is specified. However, this situation present serious difficulties towards obtaining correct statistical analysis. Biases caused by measurement errors would be difficult to access accurately without extra observations and information. One solution is to use validation data. Some examples where validation data are available can be found in [6,11,21] among others.

With the help of validation data, some statisticians developed statistical inference techniques based on surrogate data. See, e.g. [1–4,9,11,12,14–16,18,20], among others. Carroll and Wand [4] developed a semiparametric approach using the kernel regression technique for logistic measurement error models. Pepe and Fleming [12] also considered an analogous problem with surrogate a discrete random variable. It is of interest to extend it to the following partial linear model:

$$Y = X^{\tau}\beta + g(T) + e, \tag{1.1}$$

with Y measured erroneously and both explanatory variables measured exactly, where Y is a scalar response variable, X is a p-variate explanatory vector that enters the model linearly, X^{τ} is its transpose, T is another explanatory variable that enters in a nonlinear fashion, and takes values in $[0, 1]$, β is a $p \times 1$ vector of regression parameters, $g(\cdot)$ is an unknown regression function on $[0, 1]$, e is a random statistical error, and given X and T the errors $e = Y - X^{\tau}\beta - g(T)$ are assumed to be independent and identically distributed.

If the measurement error in the response is additive, the problem obviously reduces to the standard partially linear model and hence can be handled with standard methodology. Let $\tilde{Y}$ be the surrogate variable observed for the true variable Y. We consider settings where no error equation or distribution assumption of Y given $\tilde{Y}$ is specified, but some validation data are available to relate Y and $\tilde{Y}$. In such cases, the problem of estimating β and $g(\cdot)$ cannot be handled directly by (1.1) with standard methodology. To use the surrogate data $\tilde{Y}$'s, we must rewrite model (1.1) such that $\tilde{Y}$ is related to X and T. Notice that $\tilde{Y}$, X and T have useful information in predicting the unknown Y. Therefore, it is assumed that X and T are also measured in validation data set besides Y and $\tilde{Y}$. Let $Z = (\tilde{Y}, X, T)$ and $u(Z) = E[Y|Z]$. Model (1.1) can then be rewritten as

$$u(Z) = X^{\tau}\beta + g(T) + \varepsilon, \tag{1.2}$$

where $\varepsilon = e - (Y - u(Z))$. Clearly, (1.2) does not really change model (1.1). Hence, we can develop statistical inference of β and $g(\cdot)$ based on (1.2). A natural way to estimate β and $g(\cdot)$ is first to estimate $u(z)$ by kernel regression technique from the validation data and then to use the standard technique for partly linear models (see, e.g., [17]) with primary data to estimate β and $g(\cdot)$. But, such a semiparametric method would require a large validation data set, which is difficult or expensive to obtain, in order to be feasible because of the use of kernel regression with $p + 2$ explanatory variables. That is, curse of dimension will limit the usefulness of this method. This motivates us to consider this dimensional reduction model where we suppose

$$u(z) = m(\alpha^\tau z),$$

where $m(\cdot)$ is an unknown function and α is a $(p+2)\times 1$ vector of unknown parameter. α can be first estimated by sliced inverse regression (SIR) techniques (see, e.g., [5,8,10,22]). Then, we can estimate $u(z)$ by kernel regression with univariate explanatory variate with validation data. Finally, the standard techniques for partial linear models can be used to estimate β and $g(\cdot)$. It is proved that the estimator of β is asymptotically normal and the estimator of $g(\cdot)$ is weakly consistent with a optimal rate of convergence.

This paper is organized as follows. We define the estimator of β and $g(\cdot)$ in Section 2, and state the main results in Section 3. In Section 4, we present the proofs of the main results.

2. Description of methods

Let the primary data set contain N independent and identically distributed observations of $\{(\tilde{Y}_j, X_j, T_j)_{j=n+1}^{n+N}\}$. In addition to the primary data set, a validation data set containing n independent and identically distributed observations of $\{(\tilde{Y}_i, Y_i, X_i, T_i)_{i=1}^{n}\}$, which are also independent of the primary sample set, are available.

Denote $X=(X_1, X_2, \dots, X_p)^\tau$, $\mathbf{R}(Y)=(R_1(Y), \dots, R_{p+2}(Y))^\tau=(E[\tilde{Y}|Y], E[X_1|Y], \dots, E[X_p|Y], E[T|Y])^\tau$, $\Lambda_1 := Cov(\mathbf{R}(Y)) = Cov(E(Z|Y))$, $\mathbf{R}^*(Y) = (R_1^*(Y), \dots, R_{p+2}^*(Y))^\tau = (R_1(Y)f_Y(Y), \dots, R_{p+2}(Y)f_Y(Y))^\tau$, where $f_Y(y)$ is the density function of Y. Denote by Z_{ij} the jth component of Z_i for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, p+2$. Let

$$\hat{R}^*_{nj}(Y) = \frac{1}{nh_{1,n}} \sum_{i=1}^{n} Z_{ij} K_1\left(\frac{Y - Y_i}{h_{1,n}}\right), \quad j = 1, 2, \dots, p+2,$$

$$\hat{f}_n(Y) = \frac{1}{nh_{1,n}} \sum_{i=1}^{n} K_1\left(\frac{Y - Y_i}{h_{1,n}}\right),$$

where $K_1(\cdot)$ is a kernel function and $h_{1,n}$ is a bandwidth. For each fixed $b>0$, let

$$\hat{f}_{nb}(Y) = \max(\hat{f}_n(Y), b),$$

$$\hat{\mathbf{R}}_{nb}(Y) = \left(\frac{\hat{R}^*_{nj}(Y)}{\hat{f}_{nb}(Y)}\right)_{(p+2)\times 1},$$

$$\hat{\Lambda}_n = \frac{1}{n} \sum_{j=1}^{n} (\hat{\mathbf{R}}_{nb}(Y_j))(\hat{\mathbf{R}}_{nb}(Y_j))^\tau - \left(\frac{1}{n}\sum_{j=1}^{n} \hat{\mathbf{R}}_{nb}(Y_j)\right)\left(\frac{1}{n}\sum_{j=1}^{n} \hat{\mathbf{R}}_{nb}(Y_j)\right)^\tau.$$

Let α_n be the eigenvector corresponding to the maximum eigenvalue of $\hat{\Lambda}_n$. By Zhu and Fang [22], we can estimate α by α_n. Then, $u(z) = m(\alpha^\tau z)$ can be estimated by

$$\hat{u}_n(z) = \frac{\sum_{i=1}^n K_2(\frac{\hat{\alpha}_n^\tau(z-Z_i)}{h_{2,n}})Y_i}{\sum_{i=1}^n K_2(\frac{\hat{\alpha}_n^\tau(z-Z_i)}{h_{2,n}})},$$

where $h_{2,n}$ is a bandwidth and $K_2(\cdot)$ is a kernel function. Let

$$W_{Nj}(t) = \frac{K_3(\frac{t-T_j}{h_{3,N}})}{\sum_{j=n+1}^{n+N} K_3(\frac{t-T_j}{h_{3,N}})},$$

$$\hat{g}_{1,N}(t) = \sum_{j=n+1}^{n+N} W_{Nj}(t)\tilde{Y}_j,$$

$$\hat{g}_{2,N}(t) = \sum_{j=n+1}^{n+N} W_{Nj}(t)X_j,$$

$$\tilde{W}_{ni}(t) = \frac{K_4(\frac{t-T_i}{h_{4,n}})}{\sum_{i=1}^n K_4(\frac{t-T_i}{h_{4,n}})},$$

$$\tilde{g}_{1,n}(t) = \sum_{i=1}^n \tilde{W}_{ni}(t)Y_i,$$

$$\tilde{g}_{2,n}(t) = \sum_{i=1}^n \tilde{W}_{ni}(t)X_i$$

where both $K_3(\cdot)$ and $K_4(\cdot)$ are the kernel functions and both $h_{3,N}$ and $h_{4,n}$ are bandwidths. The estimator of β is defined to be the one which minimizes $\hat{S}_{n,N}(\beta)$ given by

$$\hat{S}_{n,N}(\beta) = \frac{1}{N}\sum_{j=n+1}^{n+N}(\hat{u}_n(Z_j) - X_j^\tau\beta - \hat{g}_{1,N}(T_j) + \hat{g}_{2,N}^\tau(T_j)\beta)^2 + \frac{1}{n}\sum_{i=1}^n (Y_i - X_i^\tau\beta - \tilde{g}_{1,n}(T_i) + \tilde{g}_{2,n}^\tau(T_i)\beta)^2. \quad (2.1)$$

That is, the estimator of β, say $\hat{\beta}_{n,N}$, solves the equation

$$\frac{1}{N}\sum_{j=n+1}^{n+N}[(X_j - \hat{g}_{2,N}(T_j))(\hat{u}_n(Z_j) - \hat{g}_{1,N}(T_j) - (X_j - \hat{g}_{2,N}(T_j))^\tau\beta] + \frac{1}{n}\sum_{i=1}^n (X_i - \tilde{g}_{2,n}(T_i))[Y_i - \tilde{g}_{1,n}(T_i) - (X_i - \tilde{g}_{2,n}(T_i))^\tau\beta] = 0. \quad (2.2)$$

By solving (2.2), it is obtained that

$$\hat{\beta}_{n,N} = \hat{\Sigma}_{n,N}^{-1}\hat{A}_{n,N}, \tag{2.3}$$

where

$$\hat{\Sigma}_{n,N} = \frac{1}{N}\sum_{k=n+1}^{n+N} (X_k - \hat{g}_{2,N}(T_k))(X_k - \hat{g}_{2,N}(T_k))^{\tau} + \frac{1}{n}\sum_{i=1}^{n} (X_i - \tilde{g}_{2,n}(T_i))(X_i - \tilde{g}_{2,n}(T_i))^{\tau},$$

$$\hat{A}_{n,N} = \frac{1}{N}\sum_{k=n+1}^{n+N} (X_k - \hat{g}_{2,N}(T_k))(\hat{u}_n(Z_k) - \hat{g}_{1,N}(T_k)) + \frac{1}{n}\sum_{i=1}^{n} (X_i - \tilde{g}_{2,n}(T_i))(Y_i - \tilde{g}_{1,n}(T_i)).$$

We define the estimator of $g(\cdot)$ as follows:

$$\hat{g}_{n,N}(t) = \hat{g}_{1,N}(t) - \hat{g}_{2,N}(t)\hat{\beta}_{n,N}. \tag{2.4}$$

Clearly, one could also define the estimator of β to be the one which minimizes the second term at the right-hand side of (2.1). But, this would ignore the information contained in the primary data. To make up for the loss of accuracy, one can insteadly increase n, the number of the observations of the exact data which are, however, expensive and difficult to obtain. Hence, this procedure is impractical though simple.

For the estimators $\hat{\beta}_{n,N}$ and $\hat{g}_{n,N}(t)$, we have the following theorem.

Theorem 2.1. *Under all the assumptions listed in Section* 4, *we have*

$$\sqrt{n}(\hat{\beta}_{n,N} - \beta) \xrightarrow{\mathscr{L}} N(0, \mathbf{V}),$$

where

$$\mathbf{V} = \Sigma^{-1} V_1 \Sigma^{-\tau},$$

$$V_1 = E[(u(Z) - X^{\tau}\beta - g(T))^2 (X - E[X|T])(X - E[X|T])^{\tau}] + \lambda E[(Y - X^{\tau}\beta - g(T))^2 (X - E[X|T])(X - E[X|T])^{\tau}] + \lambda\{E[(Y - E[Y|Z])^2 (X - E[X|T])(X - E[X|T])^{\tau}] + 2E[(Y - E[Y|Z])(Y - X^{\tau}\beta - g(T))(X - E[X|T])(X - E[X|T])^{\tau}]\},$$

$$\Sigma = 2E[(X - E[X|T])(X - E[X|T])^{\tau}].$$

The first term in the asymptotic covariance of $\hat{\beta}_{n,N}$ is the contribution of validation data: the Fisher information for β in validation sample by the regression relationship between $u(Z)$, and X and T. The second term is the Fisher information in the true observations by the regression relationship between Y, and X and T. The third term represents the extra cost due to estimation of the unknown $E[Y|Z]$.

Remark 2.1. The asymptotic covariance of $\hat{\beta}_{n,N}$ can be estimated consistently by

$$V_{n,N} = \Sigma_{n,N}^{-1}[V_{n,N1} + V_{n,N2}]\Sigma_{n,N}^{-\tau},$$

where

$$V_{n,N1} = \frac{1}{N}\sum_{k=n+1}^{n+N} [(\hat{u}_n(Z_k) - X_k^\tau\hat{\beta}_{n,N} - \hat{g}_{n,N}(T_k))^2(X_k - \hat{g}_{2,N}(T_k))(X_k - \hat{g}_{2,N}(T_k))^\tau],$$

$$\begin{aligned}V_{n,N2} =& \frac{1}{n}\sum_{i=1}^{n} \{(Y_i - \hat{u}_n(Z_i))^2(X_i - \hat{g}_{2,N}(T_i))(X_i - \hat{g}_{2,N}(T_i))^\tau \\ &+ (Y_i - X_i^\tau\hat{\beta}_{n,N} - \hat{g}_{n,N}(T_i))^2(X_i - \hat{g}_{2,N}(T_i))(X_i - \hat{g}_{2,N}(T_i))^\tau \\ &+ 2(Y_i - \hat{u}_n(Z_i))(Y_i - X_i^\tau\hat{\beta}_{n,N} - \hat{g}_{n,N}(T_i))(X_i - \hat{g}_{2,N}(T_i))(X_i - \hat{g}_{2,N}(T_i))^\tau\}.\end{aligned}$$

Theorem 2.2. *Under the same conditions of Theorem* 3.1, *we have*

$$\hat{g}_{n,N}(t) - g(t) = O_p((Nh_{3,N}^{\frac{3}{2}})^{-1}) + O_p((Nh_{3,N})^{-\frac{1}{2}}) + O(h_{3,N}) + O_p(N^{-\frac{1}{2}})$$

for any $t \in [0,1]$.

Corollary. *Under all the assumptions of the theorem, if* $h_{3,N} = N^{-\frac{1}{3}}$ *we have*

$$\hat{g}_{n,N}(t) - g(t) = O_p(N^{-\frac{1}{3}}).$$

Remark 2.2. The convergent rate of $\hat{g}_{n,N}(t)$ is the same as the optimal one for the corresponding nonparametric estimator of regression function (see, e.g., [19]).

3. Simulation results

Without the assumption $u(z) = m(\alpha^\tau z)$, one may define the estimator of β, say $\hat{\beta}_{n,N}^*$, to be $\hat{\beta}_{n,N}$ with $\hat{u}_n(z)$ replaced by

$$\hat{u}_n^*(z) = \frac{\sum_{i=1}^n K_2^*(\frac{z-Z_i}{h_{2,n}^*})Y_i}{\sum_{i=1}^n K_2^*(\frac{(z-Z_i)}{h_{2,n}^*})},$$

where $K_2^*(\cdot)$ is a $(p+2)$-dimensional kernel function and $h_{2,n}^*$ is a bandwidth tending to zero. As pointed in the introduction, "the curse of dimension" limits the use of this estimator.

In this section, a simulation study was carried out to compare the proposed estimator $\hat{\beta}_{n,N}$ with $\hat{\beta}_{n,N}^*$ and the naive estimator $\hat{\beta}_{\text{Naive}}$ which is defined to be $\hat{\beta}_{n,N}$, with $\hat{u}_n(Z_k)$ replaced by $\tilde{Y}_k$ for $k = n+1, n+2, \ldots, n+N$. We considered the partly linear model $Y = X^\tau \beta + g(T) + e$, where $\beta = 1.30$, $g(t) = 3.5t^2$ if $t \in [0,1]$, $g(t) = 0$ otherwise. It is assumed that X and T follow the standard normal distribution and the uniform distribution on $[0,1]$, respectively. The surrogates $\tilde{Y}$ were generated as a standard χ^2 with one degree of freedom, while e given $Z = (X, T, \tilde{Y})$ was normally distributed with mean $(\alpha^\tau Z)^2 - X^\tau \beta - g(T)$ and variance $\delta^2 = 1$. We let $\alpha = (0.752, 0.372, 0.105)^\tau$, and we estimate α using α_n given in Section 2. The simulation were run with validation data and primary data sizes of $(n,N) = (30,150)$, $(60,300)$, $(60,150)$, $(120,300)$, $(30,600)$ and $(60,1200)$, respectively. In the simulation study, $h_{1,n}$, $h_{2,n}$, $h_{2,n}^*$, $h_{3,N}$ and $h_{4,n}$ were taken to be $n^{-\frac{15}{96}}$, $n^{-\frac{2}{3}}$, $n^{-\frac{2}{9}}$, $N^{-\frac{2}{3}}$ and $n^{-\frac{1}{2}}$, and the kernel functions $K_1(\cdot), K_2(\cdot), K_2^*(\cdot), K_3(\cdot)$ and $K_4(\cdot)$ are taken to be

$$K_1(u) = \begin{cases} -\frac{15}{8}x^2 + \frac{9}{8}, & -1 \leqslant u \leqslant 1, \\ 0 & \text{otherwise}, \end{cases}$$

$$K_2(u) = \begin{cases} \frac{15}{16}(1 - 2u^2 + u^4), & -1 \leqslant u \leqslant 1, \\ 0 & \text{otherwise}, \end{cases}$$

$$K_2^*(z) = K_2(z_1)K_2(z_2)K_2(z_3)$$

and $K_3(u) = K_4(u) = \frac{1}{2}I[-1 \leqslant u \leqslant 1]$ were used to calculate $\mathbf{R}_{nb}(Y)$, $\hat{u}_n(\cdot), \hat{u}_n^*(\cdot)$ and the weights $W_{Nj}(t)$, $j = n+1, \ldots, n+N$ and $W_{ni}(t)$, $i = 1, 2, \ldots, n$, respectively. The simulation results are presented in following Tables 1–3 to compare the bias and standard deviation (SD) of $\hat{\beta}_{n,N}$ with $\hat{\beta}_{n,N}^*$ and $\hat{\beta}_{\text{Naive}}$. The bias and SD are computed from 1000 simulated data sets of each size (n,N). That is, 1000 estimates were computed to yield the results in the tables.

From the simulation results, the naive estimator has serious bias and bigger SD than the proposed estimator. Comparing the proposed estimator with $\hat{\beta}_{n,N}^*$, $\hat{\beta}_{n,N}$ outperforms $\hat{\beta}_{n,N}^*$ in terms of the bias and SD. On the other hand, the simulation study illustrates that the change of size of validation set yields bigger effect on the proposed estimators than that of the surrogate set.

4. Assumptions and proofs of theorems

The model considered here is rather complicated. In order to prove our results, we must impose assumptions for $X, T, Y, \tilde{Y}, \varepsilon, g(\cdot), u(\cdot), E[u(z)|T=t], E[X|T=t], \Sigma, \mathbf{R}$,

Table 1
$\lambda = 5$

Estimate (β)	(n, N)			
	(30, 150)		(60, 300)	
	Bias	SD	Bias	SD
$\hat{\beta}_{n,N}$	0.0897	0.7033	0.0568	0.3513
$\hat{\beta}^*_{n,N}$	0.1124	1.1405	0.1055	0.9202
$\hat{\beta}_{\text{Naive}}$	−1.1722	1.9586	−1.1281	1.5839

Table 2
$\lambda = 2.5$

Estimate (β)	(n, N)			
	(60, 150)		(120, 300)	
	Bias	SD	Bias	SD
$\hat{\beta}_{n,N}$	0.0645	0.4546	0.0351	0.2629
$\hat{\beta}^*_{n,N}$	0.1019	0.9241	0.0932	0.6810
$\hat{\beta}_{\text{Naive}}$	−1.1054	1.7017	−1.1813	1.1743

Table 3
$\lambda = 20$

Estimate (β)	(n, N)			
	(30, 600)		(60, 1200)	
	Bias	SD	Bias	SD
$\hat{\beta}_{n,N}$	0.0753	0.6058	0.0341	0.3129
$\hat{\beta}^*_{n,N}$	0.1088	0.9543	0.0816	0.7804
$\hat{\beta}_{\text{Naive}}$	−1.3052	1.5792	−1.3294	0.8563

$\mathbf{R}^*, K_i(\cdot)$ for $i = 1, 2, 3, 4$ and $h_{i,n}$ for $i = 1, 2, 4, h_{N,3}$, etc. Hence, a large numbers of assumptions are needed in the proofs of the theorems. Before stating these assumptions, we introduce the following notations.

Let $g_1(t) = E[u(Z)|T = t]$, $g_2(t) = E[X|T = t]$. Denote by $g_{2r}(\cdot)$ and X_{ir} the rth component of $g_2(\cdot)$ and X_i, $i = 1, 2, \ldots, n$; $r = 1, 2, \ldots, p$. Let $||a - b|| = \sum |a_i - b_i|$ for any vectors a and b, where a_i and b_i are the ith component of a and b, respectively.

We will prove our theorem in the following assumptions:

(A.g) $g_1(t), g_{2r}(t)$ and $g(t)$ satisfy Lipschitz condition of order 1, $r = 1, 2, \ldots, p$.

(A.r) The density of T, say $r(t)$, exists and satisfies

$$0 < \inf_{0\leqslant t\leqslant 1} r(t) \leqslant \sup_{0\leqslant t\leqslant 1} r(t) < \infty.$$

(A.X) $\sup_t E[X_{1r}^2|T=t] < \infty$ and $EX_{1r}^4 < \infty$ for $r = 1, 2, \ldots, p$.

(A.Y) $\sup_{z\in\mathscr{Z}} E[Y^2|Z=z] < \infty$.

(A.$\tilde{\text{Y}}$) $E|\tilde{Y}|^4 < \infty$.

(A.Z) The density of Z, say $f_Z(z)$, has bounded partial derivative of order one and satisfies $NP(f_Z(Z) < \eta_N) \to 0$ for some positive constant sequence $\eta_N > 0$ tending to zero.

(A.e) $E[e|Z] = 0$ and $\sup_{z\in\mathscr{Z}} E[e^2|Z=z] < \infty$.

(A.Σ) $E[(X - E[X|T])(X - E[X|T])^\tau]$ is a positive definite matrix.

(A.u) $u(\cdot)$ has bounded partial derivative of order one.

(A.K$_1$) $K_1(\cdot)$ is symmetric about 0 with bounded support $[-1, 1]$, and is a kernel function of order 4.

(A.K$_2$) $K_2(\cdot)$ is a bounded nonnegative kernel function of order one with bounded support.

(A.K$_3$K$_4$) $K_3(\cdot)$ and $K_4(\cdot)$ are bounded kernel function with bounded support.

(A.h$_{1,n}$) As $n\to\infty, h_{1,n} \sim n^{-c_1}$, $b \sim n^{-c_2}$ with positive numbers c_1 and c_2 satisfying that $\frac{1}{8} + \frac{c_2}{4} < c_1 < \frac{1}{4} - c_2$, and the notation "$\sim$" means that two quantities have the same coverage order.

(A.h$_{3,N}$h$_{2,n}$) $\frac{h_{2,n}}{h_{3,N}}$ is bounded away from zero and infinity.

(A.h$_{2,n}$) $nh_{2,n}^{\frac{3}{2}}\eta_N \to 0$ and $\frac{nh_{2,n}^3}{\eta_N^2} \to 0$.

(A.h$_{3,N}$) $Nh_{3,N}\eta_N \to \infty$.

(A.h$_{4n}$) $nh_{4n} \to \infty$ and $nh_{4n}^4 \to 0$.

(A.Nn) $\frac{N}{n} \to \lambda$, where λ is a nonnegative constant.

(A.R*) $R_i^*(y)$ for $i = 1, 2, \ldots, p+2$, and $f_Y(y)$ are 3-times differentiable and their third derivatives satisfy the following conditions: there exists a neighborhood of the origin, say U, and a constant $c > 0$ such that, for any $u \in U$

$$|f_Y^{(3)}(y+u) - f_Y^{(3)}(u)| \leqslant c|u|,$$

$$|R_i^{*(3)}(y+u) - R_i^{*(3)}(u)| \leqslant c|u|, \quad i = 1, 2, \ldots, p+2.$$

(A.R) (i) For pair $1 \leqslant i, l \leqslant p+2$ and for any $u \in U$

$$|R_i(y+u)R_l(y+u) - R_i(y)R_l(y)| \leqslant C|u|.$$

(ii) $\sqrt{n}ER_i(Y)R_l(Y)I[f_Y(Y) < b] = o(1)$ as $n\to\infty$, for $1 \leqslant i, l \leqslant d$, where $I[\cdot]$ is the indicator function and b satisfies (A.h$_{1,n}$).

In what follows, we will first prove Theorem 2.1 based on the above assumptions.

Remark 4.1. Assumptions (A.R*), (A.R), (A.X), (A.$\tilde{Y}$), (A.K_1), (A.h_{1n}) are used by Zhu and Fang [22] to prove the root n consistency of α_n.

Note that

$$\hat{\beta}_{n,N} - \beta = \hat{\Sigma}_{n,N}^{-1}(\tilde{A}_{n,N}(\beta) + \tilde{A}_n(\beta)), \tag{4.1}$$

where

$$\tilde{A}_{n,N}(\beta) = \frac{1}{N}\sum_{k=n+1}^{n+N} (X_k - \hat{g}_{2,N}(T_k))(\hat{u}_n(Z_k) - \hat{g}_{1,N}(T_k) - (X_k - \hat{g}_{2,N}(T_k))^\tau \beta),$$

$$\tilde{A}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} (X_i - \tilde{g}_{2,n}(T_i))(Y_i - \tilde{g}_{1,n}(T_i) - (X_i - \tilde{g}_{2,n}(T_i))^\tau \beta).$$

To prove theorems, we first prove the following lemmas.

Lemma 4.1. *Under conditions of Theorem* 2.1, *we have*

$$\tilde{A}_{n,N}(\beta) = \frac{1}{N}\sum_{k=n+1}^{n+N}(X_k - E[X_k|T_k])\varepsilon_k + \frac{1}{Nn}\sum_{k=n+1}^{n+N}\sum_{i=1}^{n}\frac{(X_k - E[X_k||T_k])(Y_i - u(Z_k))K_2(\frac{\alpha^\tau(Z_i - Z_k)}{h_{2,n}})}{h_{2,n}f_Z(Z_k)} + o_p(N^{-\frac{1}{2}}).$$

Proof. Let

$$A_{N,1} = \frac{1}{N}\sum_{k=n+1}^{n+N} (X_k - \hat{g}_{2,N}(T_k))(\hat{u}_n(Z_k) - u(Z_k)),$$

$$A_{N,2} = \frac{1}{N}\sum_{k=n+1}^{n+N} (X_k - \hat{g}_{2,N}(T_k))\varepsilon_k.$$

Notice that

$$\varepsilon_k = e_k - (Y_k - u(Z_k)) = u(Z_k) - X_k^\tau\beta - g(T_k).$$

We have

$$\tilde{A}_{n,N}(\beta) = A_{N,1} + A_{N,2} + r_{N,1} + r_{N,2}, \tag{4.2}$$

where

$$r_{N,1} = \frac{1}{N}\sum_{k=n+1}^{n+N} (X_k - \hat{g}_{2,N}(T_k))\left[g(T_k) - \sum_{j=n+1}^{n+N} W_{Nj}(T_k)(u(Z_j) - X_j^\tau\beta)\right],$$

$$r_{N,2} = -\frac{1}{N}\sum_{k=n+1}^{n+N} (X_k - \hat{g}_{2,N}(T_k))\sum_{j=n+1}^{n+N} W_{Nj}(T_k)(\hat{u}_n(Z_j) - u(Z_j)).$$

In the appendix, we shall prove

$$\sqrt{N} r_{N,i} \xrightarrow{\text{p}} 0, \quad i = 1, 2 \tag{4.3}$$

and

$$A_{N,1} = A_{N,3} + o_p(N^{-\frac{1}{2}}), \tag{4.4}$$

where

$$A_{N,3} = \frac{1}{N} \sum_{k=n+1}^{n+N} (X_k - \hat{g}_{2,N}(T_k))(\tilde{u}_n(Z_k) - u(Z_k))$$

with

$$\tilde{u}_n(Z_k) = \frac{\sum_{i=1}^{n} Y_i K_2(\frac{\alpha^\tau (Z_i - Z_k)}{h_{2,n}})}{\sum_{i=1}^{n} K_2(\frac{\alpha^\tau (Z_i - Z_k)}{h_{2,n}})}.$$

Clearly,

$$\begin{aligned} A_{N,3} =& \frac{1}{N} \sum_{k=n+1}^{n+N} (X_k - E[X_k | T_k])(\tilde{u}_n(Z_k) - u(Z_k)) \\ &+ \frac{1}{N} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} W_{Nj}(T_k)(E[X_j | T_j] - X_j)(\tilde{u}_n(Z_k) - u(Z_k)) \\ &+ \frac{1}{N} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} W_{Nj}(T_k)(g_2(T_k) - g_2(T_j))(\tilde{u}_n(Z_k) - u(Z_k)) \\ :=& A_{N,4} + r_{N,3} + r_{N,4}. \end{aligned} \tag{4.5}$$

By the appendix, we have

$$r_{N,i} = o_p(N^{-\frac{1}{2}}), \quad i = 3, 4. \tag{4.6}$$

Let

$$A_{N,5} = \frac{1}{nN} \sum_{k=n+1}^{n+N} \sum_{i=1}^{n} \frac{(X_k - E[X_k | T_k])(Y_i - u(Z_k)) K_2(\frac{Z_i - Z_k}{h_{2,n}})}{h_{2,n} f_Z(Z_k)}.$$

Notice that

$$A_{N,5} = \frac{1}{N} \sum_{k=n+1}^{n+N} (X_k - E[X_k | T_k]) \frac{\tilde{l}(Z_k) - \tilde{f}_Z(Z_k) u(Z_k)}{f_Z(Z_k)},$$

where

$$\tilde{l}(Z_k) = \frac{1}{n h_{2,n}} \sum_{i=1}^{n} Y_i K_2\left(\frac{\alpha^\tau (Z_i - Z_k)}{h_{2,n}}\right),$$

$$\tilde{f}_Z(Z_k) = \frac{1}{nh_{2,n}} \sum_{i=1}^{N} K_2\left(\frac{\alpha^\tau(Z_i - Z_k)}{h_{2,n}}\right).$$

Hence,

$$A_{n,4} = A_{N,5} + r_{N,5}, \tag{4.7}$$

where

$$r_{N,5} = \frac{1}{N} \sum_{k=n+1}^{n+N} (X_k - E[X_k|T_k]) \frac{(\tilde{l}(Z_k) - f_Z(Z_k)u(Z_k))(f_Z(Z_k) - \tilde{f}_Z(Z_k))}{f_Z(Z_k)\tilde{f}_Z(Z_k)}.$$

In the appendix, we shall prove

$$r_{N,5} = o_p(N^{-\frac{1}{2}}). \tag{4.8}$$

Observe that

$$\begin{aligned} A_{N,2} &= \frac{1}{N} \sum_{k=n+1}^{n+N} (X_k - E[X_k|T_k])\varepsilon_k \\ &\quad + \frac{1}{N} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} W_{Nj}(T_k)(g_2(T_k) - g_2(T_j))\varepsilon_k \\ &\quad - \frac{1}{N} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} W_{Nj}(T_k)(X_j - E[X_j|T_j])\varepsilon_k \\ &:= A_{N,6} + r_{N,6} + r_{N,7}. \end{aligned} \tag{4.9}$$

By the appendix, we have

$$r_{N,i} = o_p(N^{-\frac{1}{2}}), \quad i = 6, 7. \tag{4.10}$$

Combining (4.2)–(4.10), we get

$$\tilde{A}_{n,N}(\beta) = A_{N,6} + A_{N,5} + o_p(N^{-\frac{1}{2}}). \tag{4.11}$$

This proves Lemma 4.1. $\square$

Similarly, we can prove the following Lemma 4.2.

Lemma 4.2. *Under assumptions* (A.K$_3$K$_4$), (A.g), (A.r), (A.X) *and* (A.Y), *as* $nh_{4,n} \to \infty$ *we have*

$$\tilde{A}_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} (X_i - E[X_i|T_i])\varepsilon_i + o_p(n^{-\frac{1}{2}}).$$

Lemma 4.3. *Under conditions of Theorem* 2.1, *we have*

$$\sqrt{N}(\tilde{A}_n(\beta) + \tilde{A}_{n,N}(\beta)) \xrightarrow{\mathscr{L}} N(0, \tilde{V})$$

where

$$\begin{aligned}\tilde{V} = & E[(u(Z) - X^\tau\beta - g(T))^2(X - E[X|T])(X - E[X|T])^\tau] \\ & + \lambda\{E[(Y - E[Y|Z])]^2(X - E[X|T])(X - E[X|T])^\tau] \\ & + E[(Y - X^\tau\beta - g(T))^2(X - E[X|T])(X - E[X|T])^\tau] \\ & + 2E[(Y - E[Y|Z])(Y - X^\tau\beta - g(T))(X - E[X|T])(X - E[X|T])^\tau]\}.\end{aligned}$$

Proof of Lemma 3.3. Let $U_k = (\varepsilon_k, Z_k), V_i = (Y_i, Z_i),\ i = 1, 2, \dots, n; \quad k = n+1, \dots, n+N$, and

$$\begin{aligned}\Psi_n(U_k, V_i; h_{2,n}) = & (X_i - E[X_i|T_i])\varepsilon_i + (X_k - E[X_k|T_k])\varepsilon_k \\ & + \frac{(X_k - E[X_k|T_k])(Y_i - u(Z_k))K_2(\frac{\alpha^\tau(Z_i - Z_k)}{h_{2,n}})}{h_{2,n}f_Z(Z_k)}.\end{aligned}$$

Then, by Lemmas 4.1 and 4.2 we have

$$\sqrt{N}(\tilde{A}_n(\beta) + \tilde{A}_{n,N}(\beta)) = \frac{1}{n\sqrt{N}} \sum_{i=1}^{n} \sum_{k=n+1}^{n+N} \Psi_n(U_k, V_i; h_{2,n}) + o_p(1). \tag{4.12}$$

For any p-dimensional vector γ, let

$$\psi_n(U, V; h_{2,n}) = \gamma^\tau \Psi_n(U, V; h_{2,n}),$$

$$U_{n,N} = \frac{1}{n\sqrt{N}} \sum_{i=1}^{n} \sum_{k=n+1}^{n+N} \psi_n(U_k, V_i; h_{2,n}).$$

Clearly, $U_{n,N}$ is a two sample statistic and

$$\begin{aligned}E[\psi_n(U, V; h_{2,n})|V] \to & \gamma^\tau(X - E[X|T])(Y - E[Y|Z]) \\ & + \gamma^\tau(X - E[X|T])\varepsilon.\end{aligned} \tag{4.13}$$

By (A.u) and (A.K$_2$), it follows that

$$\begin{aligned}E[\psi_n(U, V; h_{2,n})|U] \to & \gamma^\tau(X - E[X|T])\varepsilon \\ & + \frac{\gamma^\tau(X - E[X|T]) \int (u(z) - u(Z))K_2(\frac{\alpha^\tau(z - Z)}{h_{2,n}})\, dz}{h_{2,n}f_Z(Z)} \\ \to & \gamma^\tau(X - E[X|T])\varepsilon.\end{aligned} \tag{4.14}$$

Notice that

$$\begin{aligned}E[\psi_n(U, V; h_{2,n})] = & E\{E[\phi_n(U, V; h_{2,n})|U]\} \\ = & E\frac{\gamma^\tau(X - E[X|T]) \int (u(z) - u(Z))K_2(\frac{\alpha^\tau(z - Z)}{h_{2,n}})\, dz}{h_{2,n}f_Z(Z)}\end{aligned} \tag{4.15}$$

and

$$\begin{aligned}
&\int (u(z) - u(Z))K_2\left(\frac{\alpha^\tau(z-Z)}{h_{2,n}}\right) dz \\
&\quad = h_{2,n} \int (u(Z + h_{2,n}v) - u(Z))K_2(\alpha^\tau v)\, dv \\
&\quad = h_{2,n}^2 \sum_{i=1}^{p+2} \int \frac{\partial u(Z + \theta h_{2,n} v)}{\partial Z_i} v_i K_2(\alpha^\tau v)\, dv,
\end{aligned} \tag{4.16}$$

where Z_i and v_i are the ith component of Z and v, and $\theta = (\theta_1, \ldots, \theta_{p+2})$ with $0 < \theta_i < 1$. By (A.u), (A.K$_2$), (4.15), and (4.16), it follows that

$$|E[\psi_n(U, V; h_{2,n})]| \leqslant C h_{2,n}. \tag{4.17}$$

By conditions (A.Nn), and $nh_n^2 \to 0$ which is implied (A.h$_{2,n}$), we have

$$\sqrt{N} E\psi_n(U, V; h_{2,n}) \to 0. \tag{4.18}$$

Lemma B.1 of Sepanski and Lee [16] together with (4.13), (4.14) and (4.18) proves

$$\frac{1}{n\sqrt{N}} \sum_{i=1}^{n} \sum_{k=n+1}^{n+N} \psi_n(U_k, V_i; h_{2,n}) \xrightarrow{\mathscr{L}} N(0, V_1),$$

where

$$\begin{aligned}
V_1 = &E[(u(Z) - X^\tau\beta - g(T))^\tau \gamma^\tau (X - E[X|T])(X - E[X|T])^\tau \gamma] \\
&+ \lambda\{E[(Y - E[Y|Z])^2 \gamma^\tau (X - E[X|T])(X - E[X|T])^\tau \gamma] \\
&+ E[(Y - X^\tau\beta - g(T))^2 \gamma^\tau (X - E[X|T])(X - E[X|T])^\tau \gamma] \\
&+ 2E[(Y - E[Y|Z])(Y - X^\tau\beta - g(T))\gamma^\tau (X - E[X|T])(X - E[X|T])^\tau \gamma]\}.
\end{aligned}$$

This together with (4.12) proves Lemma 4.3 since γ is any p-dimensional vector.

Lemma 4.4. *Under assumptions* (A.g), (A.K$_3$K$_4$), (A.r) *and* (A.X), *as* $Nh_{3,N} \to \infty$ *and* $nh_{4,n} \to \infty$, *we have*

$$\hat{\Sigma}_{n,N} \xrightarrow{\text{a.s.}} \Sigma,$$

where $\Sigma = 2E[(X - E[X|T])(X - E[X|T])^\tau]$.

Proof. This proof is similar to that of Lemma A.2 of Wang [20]. □

Proof of Theorem 3.1. Notice that

$$\begin{aligned}
\sqrt{N}(\hat{\beta}_{n,N} - \beta) = &\sqrt{N}\Sigma^{-1}(\tilde{A}_n(\beta) + \tilde{A}_{n,N}(\beta)) \\
&+ \sqrt{N}(\hat{\Sigma}_{n,N}^{-1} - \Sigma^{-1})(\tilde{A}_n(\beta) + \tilde{A}_{n,N}(\beta)).
\end{aligned} \tag{4.19}$$

Eq. (4.19), Lemmas 4.3 and 4.4 together prove Theorem 3.1. □

Proof of Theorem 3.2. Notice that

$$\hat{g}_{n,N}(t) - g(t) = \hat{g}_{1,N} - g_1(t) - (\hat{g}^{\tau}_{2,N}(t) - g^{\tau}_2(t))\hat{\beta}_{n,N} + g^{\tau}_2(t)(\hat{\beta}_{n,N} - \beta). \quad (4.20)$$

Notice that

$$\begin{aligned}
&\hat{g}_{1,N}(t) - g_1(t) \\
&\quad = \sum_{j=n+1}^{n+N} W_{Nj}(t)(\hat{u}_n(Z_j) - u(Z_j)) + \sum_{j=n+1}^{n+N} W_{Nj}(t)(u(Z_j) - Y_j) \\
&\qquad + \sum_{j=n+1}^{n+N} W_{Nj}(t)(Y_j - E[Y_j|T_j]) + \sum_{j=n+1}^{n+N} W_{Nj}(t)(g_1(T_j) - g_1(t)) \\
&\quad := R_{N1} + R_{N2} + R_{N3} + R_{N4}.
\end{aligned} \quad (4.21)$$

Using the analogous arguments as before, we can prove that $R_{N1} = O_p((Nh_N^{\frac{3}{2}})^{-1}) + O_p((Nh_N)^{-\frac{1}{2}})$, $R_{Ni} = O_p((Nh_N)^{-\frac{1}{2}})$, $i = 2, 3$ and $ER_{N4} = O(h_N)$. This together with (4.21) yields that

$$\hat{g}_{1,N}(t) - g_1(t) = O_p((Nh_n^{\frac{3}{2}})^{-1}) + O_p((Nh_N)^{-\frac{1}{2}}) + O(h_N). \quad (4.22)$$

Similarly, we can prove that

$$\hat{g}_{2,N}(t) - g_2(t) = O_p((Nh_N)^{-\frac{1}{2}}) + O(h_N). \quad (4.23)$$

Theorem 3.1 implies that

$$\hat{\beta}_{n,N} - \beta = O_p(N^{-\frac{1}{2}}). \quad (4.24)$$

Eqs. (4.20)–(4.24) together prove Theorem 3.2. $\square$

Acknowledgments

This work is supported by the National Natural Scientific Foundation of China (Grant 10241001 and Key grant 10231030).

Appendix

In the appendix, we will prove (4.3), (4.4), (4.6), (4.8), (4.10) by the conditions of Lemma 4.1. To do so, the following Lemma A.1 due to Wang [20] is needed.

Lemma A.1. *Under assumptions* (A.r) *and* (A.K$_3$K$_4$), *we have as* $Nh_N \to \infty$

(a) $E[W_{Nj}(T_i)]^{\gamma} \leqslant c(N^{\gamma}h_N)^{-1}, \gamma = 2, 4, \ i, j = n+1, \ldots, n+N,$

(b) $E[W_{Nj}(t)]^{\gamma} \leqslant c(N^{\gamma}h_N)^{-1}, \gamma = 2, 4, \ i, j = n+1, \ldots, n+N.$

Proof of (4.3). Denote by $r_{N,1}^{[1]}$ and $r_{N,2}^{[r]}$ the rth component of $r_{N,1}$ and $r_{N,2}$. Then

$$\begin{aligned}
\sqrt{N} r_{N,1}^{[r]} =& \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} (X_k - E[X_k|T_k]) \left(g(T_k) - \sum_{j=n+1}^{n+N} W_{Nj}(T_k) g(T_j) \right) \\
& - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} \sum_{l=n+1}^{n+N} W_{Nj}(T_k)(g(T_k) - g(T_l))(X_j - E[X_j|T_j]) \\
& - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} W_{Nj}(T_k)(X_k - E[X_k|T_k])\varepsilon_j \\
& + \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \left[\sum_{j_1=n+1}^{n+N} \sum_{j_2=n+1}^{n+N} W_{Nj_1}(T_k) W_{Nj_2}(T_k)(g(T_k) - g(T_{j_1})) \right. \\
& \left. \times (g_{2r}(T_k) - g_{2r}(T_{j_2})) \right] \\
& - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j_1=n+1}^{n+N} \sum_{j_2=n+1}^{n+N} W_{Nj_1}(T_k) W_{Nj_2}(T_k)(g(T_k) - g(T_{j_1}))\varepsilon_{j_2} \\
& + \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j_1=n+1}^{n+N} \sum_{j_2=n+1}^{n+N} W_{Nj_1}(T_k) W_{Nj_2}(T_k)(X_{j_1} - E[X_{j_1}|T_{j_1}])\varepsilon_{j_2} \\
\triangleq & \sum_{i=1}^{6} r_{N,1i}^{[r]}, \quad r = 1, 2, \ldots, p. \qquad \text{(A.1)}
\end{aligned}$$

Employing conditions (A.g), (A.K_3K_4) and (A.r), Lemma A.1 and Dharmadhikari–Jacob (D–J) inequality (see, e.g., [13, p. 128]), we have

$$\begin{aligned}
E[r_{N,11}^{[r]}]^2 &\leqslant \frac{1}{N} \sum_{k=n+1}^{n+N} E\left\{ \left[\left(\sum_{j=n+1}^{n+N} W_{Nj}(T_k)(g(T_k) - g(T_j))^2 E[X_k^2|T_k] \right\} \right. \right. \\
&\leqslant C h_N^2 \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} E W_{Nj}^2(T_k) \leqslant C h_N \to 0. \qquad \text{(A.2)}
\end{aligned}$$

Similarly, we can prove

$$E[r_{N,1i}^{[r]}]^2 \leqslant C h_N \to 0, \quad i = 2, 5, \qquad \text{(A.3)}$$

$$E|r_{N,14}^{[r]}| \leqslant C\sqrt{N} h_N^2 \to 0. \qquad \text{(A.4)}$$

By (A.r), (A.K_3K_4), (A.e) and (A.X), we have

$$E(r_{N,13}^{[r]})^2 = \frac{1}{N} \sum_{j=n+1}^{n+N} E\left\{ \left(\sum_{k=n+1}^{n+N} W_{Nj}(T_k)(X_{kr} - E[X_{kr}|T_k]) \right)^2 E[\varepsilon_j^2|X_j, T_j] \right\}$$

$$\leqslant \frac{C}{N} \sum_{j=n+1}^{n+N} \sum_{k=n+1}^{n+N} E\{W_{Nj}^2(T_k)E[(X_{kr} - E[X_{kr}|T_k])^2|T_k]\}$$
$$\leqslant C(Nh_N)^{-1} \to 0. \tag{A.5}$$

Similarly, we can prove

$$E|r_{N,16}^{[r]}| \leqslant C(Nh_N)^{-1} \to 0. \tag{A.6}$$

Eqs. (A.1)–(A.6) together prove

$$\sqrt{N} r_{N,1} \xrightarrow{\text{p}} 0. \tag{A.7}$$

It is clear that

$$\begin{aligned}\sqrt{N} r_{N,2} = & -\frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} W_{Nj}(T_k)(X_k - E[X_k|T_k])(\hat{u}_n(Z_j) - u(Z_j)) \\ & - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j_1=n+1}^{n+N} \sum_{j_2=n+1}^{n+N} W_{Nj_1}(T_k) W_{Nj_2}(T_k)(g_2(T_k) - g_2(T_{j_1})) \\ & \times (\hat{u}_n(Z_{j_2}) - u(Z_{j_2})) \\ & - \frac{1}{\sqrt{N}} \sum_{k=n+1}^{n+N} \sum_{j_1=n+1}^{n+N} \sum_{j_2=n+1}^{n+N} W_{Nj_1}(T_k) W_{Nj_2}(T_k)(E[X_{j_1}|T_{j_1}] - X_{j_1}) \\ & \times (\hat{u}_n(Z_{j_2}) - u(Z_{j_2})) \\ := & \, r_{N,21} + r_{N,22} + r_{N,23}. \end{aligned} \tag{A.8}$$

Recalling the definition of $\hat{u}_n(\cdot)$, $r_{N,21}$ can be represented as

$$r_{N,21} = r_{N,211} + r_{N,212}, \tag{A.9}$$

where $r_{N,211}$ and $r_{N,212}$ are $r_{N,21}$ with $u(\cdot)$ replaced by $\tilde{u}_n(\cdot)$ and $r_{N,21}$ with $\hat{u}_n(\cdot)$ replaced by $\tilde{u}_n(\cdot)$, respectively.

By Theorem 2 of Zhu and Fang [22], we have

$$\hat{\alpha}_n - \alpha = O_p(n^{-\frac{1}{2}}). \tag{A.10}$$

Hence, the same arguments as those used in the proof of Theorem 3.3 of Härdle and Stoke (1989) can be applied to the proof of the following (A.11).

$$P\left(\sup_z |\hat{f}_Z(Z_k) - \tilde{f}_Z(Z_k)| > \eta_N\right) \to 0 \tag{A.11}$$

as $\eta_N N^{\frac{2}{5}} \to \infty$, which is implied by (A.$h_{3,N}$). Employing (A.10) and (A.11), standard arguments can be used to prove

$$r_{N,211} \xrightarrow{\text{p}} 0. \tag{A.12}$$

Using arguments similar to Wang [20], it can be shown that

$$r_{N,212} \xrightarrow{p} 0. \tag{A.13}$$

Combining (A.9), (A.10), (A.12) and (A.13), it follows that

$$r_{N,21} \xrightarrow{p} 0. \tag{A.14}$$

Similarly, we can prove

$$r_{N,2i} \xrightarrow{p} 0, \quad i = 2, 3. \tag{A.15}$$

From (A.8), (A.14) and (A.15), we get

$$\sqrt{N} r_{N,2} \xrightarrow{p} 0. \tag{A.16}$$

Eqs. (A.7) and (A.16) together prove (3.2). □

Proof of (4.4). Notice that

$$A_{N,1} - A_{N,3} = \frac{1}{N} \sum_{k=n+1}^{n+N} (X_k - \hat{g}_{2,N}(T_k))(\hat{u}_n(Z_k) - \tilde{u}_n(Z_k)). \tag{A.17}$$

Using the inequality used in the proof of Theorem 3.3 of Härdle and Stoke (1989) and (A.10), the similar arguments to those used in the proof of (A.16) can be applied to the proof of (4.4). The detail is omitted. □

Proof of (4.6). Using similar arguments to Wang [20], we can prove (4.6). □

Proof of (4.8). Eq. (4.8) can be proved following the proof of (A.46) in [20] line by line. □

Proof of (4.10). By (A.g), (A.r), (A.K_3K_4), (A.e) and Lemma A.1, we have

$$\begin{aligned} E(r_{N,6})^2 &= \frac{1}{N} \sum_{k=n+1}^{n+N} E\left[\sum_{j=n+1}^{n+N} W_{Nj}(T_k)(g_2(T_k) - g_2(T_j))^2 E[\varepsilon_k^2 | T = T_k] \right\} \\ &\leqslant Ch_N^2 \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} E\left[W_{Nj}^2(T_k) \left| \frac{T_k - T_j}{h_{2,n}} \right|^2 \right] \\ &\leqslant Ch_N^2 \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} EW_{Nj}^2(T_k) \\ &\leqslant Ch_N \to 0. \end{aligned} \tag{A.18}$$

Notice that (C.e) (i) implies $E[\varepsilon | X, T] = 0$, and (A.e) (ii) and (A.Y) imply that $\sup_{x \in \mathscr{X}, 0 \leqslant t \leqslant 1} E[\varepsilon^2 | X = x, T = t] < +\infty$. Hence, by D–J inequality, we have

$$E(r_{N,7})^2 \leqslant \frac{1}{N} \left\{ \left(\sum_{j=n+1}^{n+N} E[W_{Nj}(T_k)(X_{jr} - E[X_{jr} | T_j])]^2 E[\varepsilon_k^2 | X_{kr}, T_{kr}] \right\}$$

$$\leqslant \frac{c}{N} \sum_{k=n+1}^{n+N} \sum_{j=n+1}^{n+N} E\{W_{Nj}^2(T_k)E[(X_j - E[X_j|T_j])^2|T_j]\}$$
$$\leqslant C(Nh_N)^{-1}.$$

Eqs. (A.18) and (A.19) together prove (4.10). $\square$

References

[1] J. Bound, A.B. Krueger, The extent of measurement error in longitudinal earnings data: do two wrongs make a right?, J. Labor Econometrics 9 (1989) 1–24.

[2] R.J. Carroll, R.K. Knickerbocker, C.Y. Wang, Dimension reduction in a semiparametric regression model with errors in covariables, Ann. Statist. 23 (1995) 161–181.

[3] R.J. Carroll, L.A. Stefanski, Approximate quasilikelihood estimation in models with surrogate predictors, J. Amer. Statist. Assoc. 85 (1990) 652–663.

[4] R.J. Carroll, M.P. Wand, Semiparametric estimation in logistic measure error models, J. Roy. Statist. Soc. Ser. B 53 (1991) 652–663.

[5] N. Duan, K.C. Li, Slicing regression: a link free regression method, Ann. Statist. 19 (1991) 505–530.

[6] G. Duncan, D. Hill, An investigations of the extent and consequences of measurement error in labor-economics survey data, J. Labor Economics 3 (1985) 508–532.

[7] W.A. Fuller, Measurement Error Models, Wiley, New York, 1987.

[8] T. Hsing, R.J. Carroll, Asymptotic theory for sliced inverse regression, Ann. Statist. 20 (1992) 1040–1061.

[9] L.-F. Lee, J.H. Sepanski, Estimation of linear and nonlinear error-in-variable models using validation data, J. Amer. Statist. Assoc. 90 (1995) 130–140.

[10] K.C. Li, Sliced inverse regression for dimension reduction (with discussion), J. Amer. Statist. Assoc. 86 (1991) 337–342.

[11] M.S. Pepe, Inference using surrogate outcome data and a validation sample, Biometrika 79 (1992) 355–365.

[12] M.S. Pepe, T.R. Fleming, A general nonparametric method for dealing with errors in missing or surrogate covariable data, J. Amer. Statist. Assoc. 86 (1991) 108–113.

[13] B.L.S.P. Rao, Nonparametric Functional Estimation, Academic Press, Inc., London, 1983.

[14] J.H. Sepanski, R.J. Carroll, Semiparametric quasilikelihood and variance function estimation in measurement error models, J. Econometrics 58 (1993) 223–256.

[15] J.H. Sepanski, R. Kinickerbocker, R.J. Carroll, A semiparametric correction for attenuation, J. Amer. Statist. Assoc. 89 (1994) 1366–1373.

[16] J.H. Sepanski, L.F. Lee, Semiparametric estimation of nonlinear error-in-variables models with validation study, J. Nonparam. Statistics 4 (1995) 365–394.

[17] P. Speckman, Kernel smoothing in partial linear models, J. Roy. Statist. Soc. Ser. B 50 (1988) 413–436.

[18] L.A. Stefanski, R.J. Carroll, Conditional scores and optimal scores for generalized linear measurement error models, Biometrika 74 (1987) 703–716.

[19] C.J. Stone, Optimal rates of convergence for nonparametric estimators, Ann. Statist. 8 (1980) 1348–1360.

[20] Q.H. Wang, Estimation of partial linear error-in-variables model with validation data, J. Multivariate Anal. 69 (1999) 30–64.

[21] J. Wittes, E. Lakatos, J. Probstfied, Surrogate endpoints in clinical trails: cardiovascular diseases, Statist. Med. 8 (1989) 415–425.

[22] L.X. Zhu, K.T. Fang, Asymptotics for kernel estimator of sliced inverse regression, Ann. Statist. 24 (1996) 1053–1068.